

Effects of spectral alternation on the intelligibility of words and sentences

JAMES A. BASHFORD, JR., and RICHARD M. WARREN
University of Wisconsin—Milwaukee, Milwaukee, Wisconsin

In this study of a novel speech interference paradigm, it was found that periodic alternation between nonoverlapping filtering conditions (1100 Hz low-pass and 1700 Hz high-pass) significantly reduced intelligibility. Thus, although repetition accuracy for key words within sentences was approximately 98% with continuous low-pass or high-pass filtering alone, listeners averaged only 48% correct when frequency ranges alternated at a rate of 2.67 Hz. Interestingly, intelligibility improved by about 39% when low-pass noise was mixed with segments of high-pass speech and high-pass noise was mixed with segments of low-pass speech. A smaller increase due to added noise was also observed with spectrally alternated word lists. The interference produced by spectral alternation appears to be related to other verbal and musical phenomena, including stream segregation. The results from various control conditions indicate that the enhanced intelligibility produced by added noise involved phonemic restorations that occurred within each speech band, as well as a spectral completion effect that acted across spectral bands.

When alternate notes of a familiar melody are transposed, so as to be played in separate registers by a single instrument, the original melody may become unrecognizable (Deutsch, 1972; Dowling & Hollombe, 1977). Under these conditions, a listener may hear two perceptually distinct and unfamiliar melodies, each composed of notes from only one of the two registers. Conversely, two different familiar melodies whose notes are interleaved and played by a single instrument may be heard as an unrecognizable pattern, unless the interleaved notes are played in separate registers. In the latter case, the single instrument may appear to produce the two melodies simultaneously (Bukofzer, 1947, p. 304; Dowling, 1973; Ortmann, 1926, p. 23; Piston, 1947, p. 23). This dominance of spectral over temporal contiguity is known to musicians as a "compound melodic line" or "implied polyphony," and has been used extensively by Baroque composers such as Bach. When studied experimentally it has been referred to as the "trill" effect (Miller & Heise, 1950), "fission" (van Noorden, 1975), "primary auditory stream segregation" (Bregman & Campbell, 1971), or simply "streaming."

This process of spectral segregation, which can override temporal contiguity in the perception of auditory sequences, appears to be related to selective attention. Both interleaved melodies and intermixed verbal messages can be disentangled more readily when the components of paired signals inhabit distinctly different frequency

regions. Thus, a listener can more easily comprehend simultaneous verbal messages when the competing speech signals are produced by speakers whose voices have characteristically different ranges of fundamental voicing frequency and different long-term spectra (Broadbent, 1958).

We reasoned that manipulations of the type that would be expected to produce spectral segregation and facilitate the resolution of competing signals might disrupt processing when applied to a single verbal message. Specifically, we predicted that an effect analogous to the perceptual splitting of a single known melody into two unknown sequences would occur when a speech signal was subjected to a periodic "spectral alternation" between nonoverlapping low-pass and high-pass filtering conditions (see Figure 1A). In pilot studies using our laboratory staff and naive listeners, we first determined low-pass and high-pass cutoff frequencies of filtered speech (sentences and discourse) that yielded approximately equal intelligibility when heard separately. We then applied periodic alternation between the low-pass and high-pass conditions and found that the speech did split into perceptually separate streams at alternation rates above about 2 Hz, with the perceptually separate bands appearing to behave like independent, interrupted speech signals. At alternation rates between 2 and 4 Hz, the intelligibility of these spectrally alternated signals was very low, a finding of interest because each speech band was highly intelligible when presented alone without interruption, so that it might be expected that the spectrally alternating version of the speech signal would carry a sufficient amount of information for high intelligibility as well.

Since this type of intermittent filtering of speech does not occur under natural listening conditions, listeners might find it difficult to treat the alternating bands (shown

This research was supported by grants from the National Institutes of Health (5 R01 NS19295) and the Air Force Office of Scientific Research (85-0260A). We wish to thank the reviewers (Arthur G. Samuel and an anonymous reviewer) for their helpful criticisms of an earlier version of this manuscript. The author's address is Department of Psychology, University of Wisconsin—Milwaukee, Milwaukee, WI 53201.

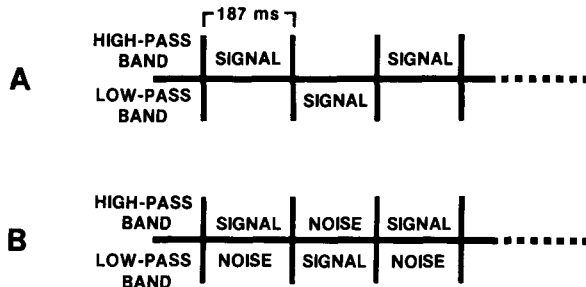


Figure 1. Alternating patterns of stimulation used in preliminary studies. The loss of intelligibility found for spectrally alternated speech (A) was largely restored when complementary noise was added to the silent gaps (B).

in Figure 1A) as components of a single signal and integrate the information provided by the spectrally disparate fragments of speech. We reasoned that the addition of complementarily filtered noise to the spectrally alternating speech (that is, mixing low-pass noise with high-pass speech and mixing high-pass noise with low-pass speech, as shown in Figure 1B) might produce a stimulus that would be treated as a single speech message subjected to interference by noises having different spectral components. Further pilot work indicated that the addition of complementary noise did indeed produce the anticipated effect. The spectrally alternated speech then appeared to be treated more like a single signal, and there was a concomitant increase in intelligibility. For example, with alternation at a rate of 2 Hz, the addition of complementary noise produced an apparent continuity of one or both speech bands. Furthermore, laboratory staff reported that the addition of noise often produced an apparent fusion of the two speech bands, causing the spectrally alternating speech signal (sentences or discourse) to appear continuous and broadband. Additional pilot work with naive listeners indicated that the addition of complementary noise also produced a reliable increase in the objective intelligibility of spectrally alternated sentences and word lists. The greatest increase in intelligibility was observed at alternation rates between 2 and 4 Hz. When noise was not added to the stimulus, those rates of alternation also produced the greatest intelligibility loss.

The loss of intelligibility produced by spectral alternation and the recovery produced by the addition of complementary noise appear to be related to interference effects previously observed under conditions of dichotic alternation, in which a broadband monaural speech signal is periodically switched between the listener's two ears, so that the message is continuously presented but never to both ears simultaneously (Cherry, 1953). As with our observations using spectral alternation, maximum interference by dichotic alternation was observed at switching rates of about 2 to 4 Hz, and interference could be reduced greatly by filling the monaural gaps produced by the interruption with broadband noise (Cherry, 1953;

Cherry & Taylor, 1954; Huggins, 1964, 1967; Schubert & Parker, 1955; Wingfield & Wheale, 1975). Perhaps the simplest kind of interrupted speech involves a diotic signal turned on and off, and earlier studies have demonstrated that speech subjected to multiple interruptions, whether periodic or aperiodic, can regain much of its intelligibility when the gaps are filled with noise (Bashford & Warren, 1979; Cherry & Wiley, 1967; Holloway, 1970; Powers & Wilcox, 1977; Verschuure & Brocaar, 1983).

The recovery of intelligibility produced by interpolated noise appears to be due to a general auditory mechanism that can restore masked portions of signals. This process, which has been called "temporal induction," uses the auditory input provided by a louder extraneous sound for abstracting appropriate neural components and reconstructing contextually appropriate obliterated segments (see Warren, 1984, for a recent review). A specialized form of temporal induction utilizes linguistic skills for reconstruction of speech. This form of induction, called "phonemic restoration" (Warren, 1970), can recreate single phonemes and syllables (Layton, 1975; Obusek & Warren, 1973; Samuel, 1981a, 1981b, 1987; Samuel & Ressler, 1986; Sasaki, 1980; Warren, 1970; Warren & Obusek, 1971; Warren & Sherman, 1974). Listeners typically are not able to distinguish a phonemically restored segment from the acoustically real speech sounds (Warren & Obusek, 1971). Similar restoration effects have also been observed with speech signals subjected to periodic interruption (Bashford & Warren, 1979; Miller & Licklider, 1950; Verschuure, 1978). When periodically spaced fragments of speech have been replaced by a louder broadband noise, illusory continuity has been found to occur through gaps ranging from 50 to 160 msec for lists of isolated monosyllables (Bashford & Warren, 1979; Miller & Licklider, 1950), and through periodic gaps as great as 300 msec for connected discourse (Bashford & Warren, 1987). Furthermore, as found for the illusory continuity of interrupted nonverbal signals (Houtgast, 1972; Warren, Obusek, & Ackroff, 1972), restoration of speech through periodic interruptions has been found to be maximal when the interpolated noise is a potential masker of the deleted segments of the signal (Bashford & Warren, 1987).

As shown in Figure 1B, when complementary noise is added to spectrally alternated speech, segments of high-pass speech alternate with segments of high-pass noise, and segments of low-pass speech alternate with segments of low-pass noise. Thus, the stimulus can be considered as two different portions of a single message that are complementary in both spectrum and time. As can be seen in Figure 1B, each spectral range of filtered speech is interrupted by noise having a similar spectrum and, thereby, is potentially subject to phonemic restoration. Our preliminary observations indicated that multiple phonemic restorations were indeed occurring under conditions of spectral alternation, and that the increase in intelligibility

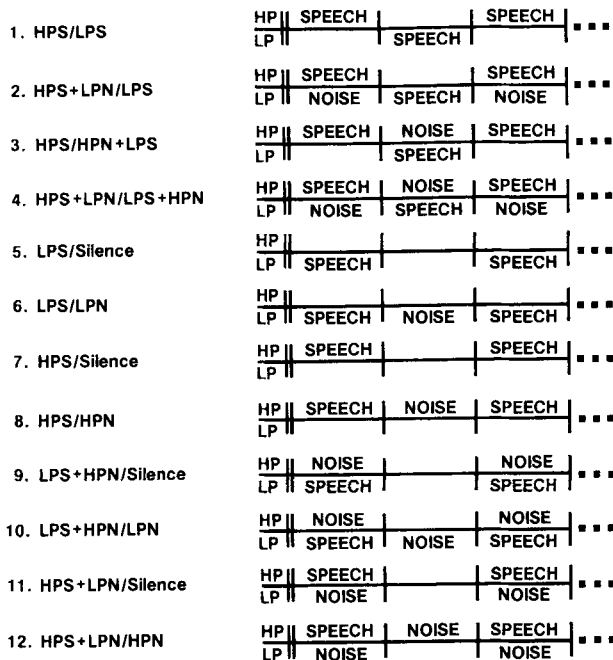


Figure 2. Diagrammatic representation of the 12 main stimulus conditions used in the experiment (187 msec on-time and off-time for all signal components). High-pass (HP) stimuli had components below 1700 Hz removed; low-pass (LP) stimuli had components above 1100 Hz removed.

produced by interpolation of noise was substantial. In order to examine the objective effects of spectral alternation and added noise in greater detail, we asked the listeners in the formal experiment described below to attempt to repeat (but not shadow) sentences and lists of isolated words subjected to spectral alternation. The interruption rate used (2.67 Hz) had been found in pilot work to produce a severe loss of intelligibility without complementary noise and a substantial recovery with noise. In addition to the main conditions of spectral alternation with and without noise, listeners were presented with 10 additional interruption conditions, as shown in Figure 2. These reduced conditions were chosen to test for two possible mechanisms of perceptual reconstruction: phonemic restoration occurring separately in each spectral region and a novel mechanism of spectral completion acting across spectral regions.

METHOD

Subjects

The 48 subjects (28 female and 20 male) were enrolled in introductory psychology courses at the University of Wisconsin-Milwaukee. They were selected from a larger pool of subjects on the basis of an audiometric screening task described under Procedure; they were either paid or given course credit for their participation.

Stimuli

The 300 isolated words used in this study were taken from the first six lists of phonetically balanced (PB) monosyllables given by the American National Standards Institute (1960). The remaining stimuli were CID (Central Institute for the Deaf) sentences (Sil-

verman & Hirsh, 1955). These sentences are highly redundant and low in abstraction, and are considered representative of everyday American speech. Each sentence contains an average of five phonetically balanced key words suggested for use in articulation testing (Davis & Silverman, 1970, pp. 481-482).

The speech stimuli were produced in a sound-attenuating chamber (IAC Series 400A) by a male speaker having a General American dialect. The output of an Altec microphone (Model 681A) was amplified by an Ampex AM-10 mixer, band-pass filtered from 100 to 8000 Hz with slopes of 48 dB/octave (Rockland Model 852 filter), and recorded on tape at 15 ips using an Ampex 440-C eight-track recorder. A panel of 3 listeners experienced in psychoacoustic research judged all words and sentences to be clearly articulated and uniform in loudness. Sound level measurement with a Brüel & Kjaer Model 2204 precision sound-level meter, operating in its impulse mode, yielded ranges of less than 6 dB in peak level across the sentences and less than 9 dB across the isolated monosyllables. Graphic level analysis (General Radio Model 1521 B graphic level recorder) yielded an average duration of 294 msec for words within the CID sentences. The isolated monosyllables, which were produced at a rate of 0.33 words/sec, averaged 583 msec in duration.

The speech recordings were subjected to on-line filtering during the experiment by being passed successively through Rockland Model 1042F and Model 852 dual hi/lo filters to produce two simultaneous versions of each speech signal, one version low-pass filtered at 1100 Hz and the other high-pass filtered at 1700 Hz (with slopes of 96 dB/octave in each case).¹ The high-pass and low-pass noise were derived from pink noise (produced by a General Radio Model 1382 generator) subjected to the same filtering as the speech signals and prerecorded on additional tracks of the eight-track tape containing the speech signals.

Apparatus

The filtered speech signals prepared as described above were passed through electronic switches (Grason-Stadler Model 1297B) to separate subchannels of an eight-channel audio mixer (Yamaha Model PM-430). The filtered noise signals were passed through equivalent electronic switches to two additional channels of the mixer. Triggering of the four electronic switches (set for 187 msec on/off with 10 msec rise/fall) was synchronized so that spectrally complementary bands of speech and noise were delivered simultaneously to the mixer. The spectrally alternating mixtures of speech and noise (high-pass speech plus low-pass noise alternating with low-pass speech plus high-pass noise) were passed from one master channel of the Yamaha mixer to an impedance-matching transformer (Grason-Stadler Model E10589A) and transduced diotically through a matched pair of Telephonics TDH-49 headphones mounted in MX 41/AR cushions. For the setting of presentation levels, the two speech bands were delivered simultaneously to the headphones at the same relative amplitudes as in the original broadband signals, and the combined signal was adjusted to produce a peak level of 80 dBA SPL. A corresponding procedure was followed in adjusting the average noise intensity to 80 dBA SPL with both noise bands present. (Neither the two speech bands nor the two noise bands were ever presented simultaneously during the experiment.)

With the exception of the Ampex recorder, all of the apparatus was located inside the audiometric chamber occupied by the subject and the experimenter, and was under the direct control of the experimenter. The recorder was operated from within the test chamber via its remote transport and cueing modules. By activating the appropriate channels of the mixer, the experimenter could deliver to the subject's headphones any of the desired combinations of filtered speech and filtered noise shown in Figure 2.

Procedure

Audiometric screening. At least 1 day prior to participation in the formal experiment, the subjects were screened individually in

an IAC single-walled sound-attenuating chamber, using a Békésy-type tracking procedure for a diotically presented sinusoidal tone swept from 125 Hz to 8 kHz in four alternately ascending and descending sweeps of one octave per minute. The subjects produced their threshold tracings by pressing and releasing a remote control switch for the audiometer (Grason-Stadler Model E-800), which produced a decrease or increase of tonal intensity at a rate of 2.5 dB/sec. Subjects whose tracings were 20 dB above the normal threshold at any frequency, or whose tracings differed across sweeps by more than 15 dB at any frequency, were not included in the formal experiment. Under these criteria (which excluded not only subjects with hearing impairments, but also those who did not follow the standard audiometric instructions), approximately 50% of the subjects qualified for further participation in the study.

Formal testing. The 36 subjects assigned to the main part of this experiment attempted to repeat the CID sentences and then the PB words presented under the 12 conditions listed in Figure 2. These conditions included both the full experimental stimulus, containing alternating mixtures of complementary speech and noise (Stimulus 4), and all speech-containing subcomponents of that stimulus. The use of 12 stimulus conditions made it necessary to construct 12 lists of CID sentences from the original 10 published lists. This was accomplished by selecting 96 of the 100 sentences and assigning them to 12 lists, with each list composed of 8 sentences containing a total of 40 key words to be scored. The experimental conditions were assigned to sentence lists in a pseudorandom fashion for each subject, with the restriction that each list be presented in each condition three times across subjects. The sentences were presented in eight blocks of 12, with each block containing one sentence from each list. Thus, each condition was presented once in each block. The order of condition presentation within blocks was pseudorandomized for each subject, with the restriction that no condition appear successively across blocks.

Following the sentence-repetition phase of the study, the subjects were asked to repeat the PB word stimuli. The 300 PB words were divided into 60 sets of 5 words each. These 60 word-sets were presented in five blocks of 12, with each experimental condition assigned to one word-set within each block. The order of condition presentation within blocks of word-sets was pseudorandomized for each subject, with the restriction that no condition appear successively across blocks for any subject, and that each condition was paired with each word-set three times across subjects. As was the case in the sentence-repetition phase of the study, the subjects were given as much time as they needed to respond to each stimulus, and their repetition responses were recorded to permit verification of the experimenter's initial scoring.

In order to obtain a baseline measure of the effects of filtering alone upon intelligibility, an additional 12 audiometrically screened subjects were asked to repeat the sentences and the PB word-sets when the stimuli were uninterrupted and presented either in high-pass or low-pass filtered form at the same intensity level employed in the main experiment. For each subject, half of the sentence sets and half of the PB word-sets were high-pass filtered and half were low-pass filtered. The assignment of filtering conditions to sentence and word-sets was pseudorandomized for each subject, with the restriction that each set be presented in both filtering conditions six times across subjects.

RESULTS

Baseline Conditions

Table 1 presents the repetition accuracy (percent correct) in the baseline condition (no interruption) for the 240 key words within sentences and for the 150 isolated monosyllables in each of the two filtering conditions (continuous high-pass speech and continuous low-pass speech).

Table 1
Repetition Accuracy for the Key Words in CID Sentences
and for the Monosyllabic Word Sets

Condition	Sentences		Monosyllables	
	% Correct	SE	% Correct	SE
Baseline Conditions				
Continuous HPS	98.5	0.28	79.8	0.02
Continuous LPS	98.2	0.29	59.6	0.02
Experimental Conditions				
1. HPS/LPS	48.0	2.98	17.6	1.41
2. HPS+LPN/LPS	51.9	2.64	20.0	1.78
3. HPS/HPN+LPS	76.9	1.90	30.0	2.09
4. HPS+LPN/LPS+HPN	86.7	1.63	28.7	1.47
5. LPS/Silence	13.0	1.93	3.7	0.73
6. LPS/LPN	19.9	2.25	2.9	0.49
7. HPS/Silence	19.2	2.02	10.1	1.14
8. HPS/HPN	31.6	2.16	9.6	1.03
9. LPS+HPN/Silence	18.0	2.19	3.6	0.71
10. LPS+HPN/LPN	18.1	2.14	3.0	0.68
11. HPS+LPN/Silence	15.7	1.35	5.9	0.99
12. HPS+LPN/HPN	9.0	1.60	5.9	0.97

Note—The four stimulus components were high-pass speech (HPS), low-pass speech (LPS), high-pass noise (HPN), and low-pass noise (LPN). CID = Central Institute for the Deaf (Silverman & Hirsh, 1955).

Sentence intelligibility was nearly perfect (scores of over 98%) for both filtering conditions. Accuracy scores for the monosyllables were considerably lower, and unlike performance with sentences, did show a difference between high-pass and low-pass conditions [$t(10) = 6.83$, $p < .001$].²

Experimental Conditions

Repetition accuracy scored as percent correct for key words in sentences and isolated monosyllables in the 12 conditions of the main experiment are listed under Conditions 1 through 12 in Table 1. The data for sentences and words were subjected to separate speech \times noise \times subjects analyses of variance. It should be remembered that in the factorial combination of speech conditions (low-pass speech only, high-pass speech only, and alternating low- and high-pass speech) and noise conditions (no noise, low-pass noise only, high-pass noise only, and alternating low- and high-pass noise), only speech and noise with complementary band spectra were presented simultaneously to the subject, and speech and noise bands with corresponding spectra were presented only alternately.

In the analysis of data for both sentences and words, conservative F ratios ($df = 1,35$) for noise ($F = 40.3$, $F = 9.65$), speech ($F = 654.95$, $F = 350.5$), and the interaction of noise and speech ($F = 49.5$, $F = 12.67$) were significant at the .005 level or better. The results of subsequent pairwise comparisons (Newman-Keuls tests) between the various experimental conditions are discussed in context below.³

Conditions 1-4, which presented both speech bands, examined the intelligibility of spectrally alternated speech with and without the addition of filtered noise(s). As expected on the basis of our pilot observations, Condition 1, which presented subjects with speech subjected to spec-

tral alternation without added complementary noise, did produce a large reduction in intelligibility. Repetition accuracy for sentences in that condition was about 50% less than that observed in the baseline conditions for either the low-pass or the high-pass filtered sentences presented without interruption. Similarly, accuracy for isolated monosyllables in Condition 1 was about 42% lower than for the uninterrupted low-pass band and about 62% lower than for the high-pass band of the same word lists as presented in the baseline conditions. Thus, the results obtained for Condition 1 indicate that a substantial amount of potentially useful information (corresponding to about 50% intelligibility for both sentences and word lists) was available but perceptually inaccessible to the subjects when alternating spectral gaps in the speech signals were filled with silence. Conditions 2-4, discussed below, examined the extent to which that inaccessible information could be perceptually recovered when spectral gaps were filled with appropriately filtered noise.

When gaps in only the low-pass band of the spectrally alternating speech were filled with low-pass noise (Condition 2), there was no reliable increase in repetition accuracy for either sentences or words. However, the addition of high-pass noise in Condition 3 did improve the intelligibility of both types of speech ($p < .01$). Repetition accuracy increased by 29% for sentences and by 11% for word lists. Thus, a substantial amount of verbal information that was rendered nonfunctional through spectral alternation (Condition 1) became functional when interruptions in the high-pass band of the speech signal were filled with noise having the same spectral limits. The ineffectiveness of the low-pass noise in Condition 2 was probably due to the relatively lower intelligibility of the low-pass speech band, as was found with isolated monosyllables in the baseline conditions (as will be discussed, further evidence of unequal passband intelligibility and corresponding differential noise effects were found in Conditions 5-12).

Although only the high-pass noise produced an independent enhancement of intelligibility under spectral alternation, the further addition of the low-pass noise in Condition 4 did produce an additional increase in repetition accuracy for sentences ($p < .01$). Thus, repetition performance increased from a level of 48% accuracy with no noise present to about 87% accuracy when gaps in both the low- and high-frequency regions of the sentences were filled with noise. This dramatic increase in intelligibility is consistent with our original hypothesis that the addition of complementary noise would counteract segregation of the two speech bands and facilitate processing of temporally adjacent (but spectrally disparate) speech fragments as remnants of a single signal subjected to partial masking in different frequency regions over time. As mentioned in the introduction, our pilot observations had indicated that phonemic restoration occurring in one or both speech bands was one mechanism contributing to the access of information otherwise lost through spectral alternation. Conditions 5-12, which presented single speech

bands with all component combinations of interpolated silence and added noise, were included in part to permit estimation of the contribution of phonemic restoration occurring separately in each frequency region. By determining the role of phonemic restorations, it is possible to assess the extent of another form of restoration, "spectral completion," acting across spectral regions.

In Conditions 5-8, single speech bands without complementary noise were periodically interrupted by silence (Conditions 5 and 7) or by interpolated noise having the same spectral limits as the speech (Conditions 6 and 8). As found previously for periodically interrupted lists of unfiltered monosyllables (Dirks & Bower, 1970; Miller & Licklider, 1950), interpolation of spectrally matched noise in the present experiment did not improve repetition accuracy for either the low-pass or the high-pass word list. In contrast, and also in keeping with previous studies using sentences as interrupted stimuli (Bashford & Warren, 1979; Powers & Wilcox, 1977; Verschuure & Brocaar, 1983), interpolation of spectrally matched noise in the present study did enhance repetition accuracy for filtered sentences, but only in the case of high-pass filtered sentences interrupted by high-pass noise (Condition 8) was the increase (about 12%) significant ($p < .01$). This asymmetry in the effectiveness of the two noise bands on accuracy scores for sentences paralleled their relative effects when both speech bands were present under spectral alternation (Condition 2 vs. Condition 3).

Conditions 9-12 were identical to Conditions 5-8, except that spectrally complementary noise was mixed with the filtered speech segments. Comparison of corresponding conditions indicates that, with only one speech band present, the only effect of added simultaneous noise was a reduction in repetition accuracy that occurred when low-pass noise was mixed with fragments of high-pass monosyllables interrupted by silence (Condition 7 vs. Condition 11; $p < .05$) and with fragments of high-pass sentences interrupted by high-pass noise (Condition 8 vs. Condition 12; $p < .01$).

Thus, in conditions that presented only one speech band, the only reliable effects of added noise were an enhancement of intelligibility for high-pass sentences produced by interpolated high-pass noise and a reduction of intelligibility for both the high-pass sentences and monosyllables produced by simultaneous low-pass noise. It is of interest that the addition of low-pass noise to the fragments of high-pass sentences actually reversed the otherwise beneficial effect of interpolated high-pass noise (compare Condition 7 with Conditions 8 and 12). As we shall see, the effects of noise observed in this study can be explained by the operation of two pairs of opposing processes: spectrally limited phonemic restorations and "spectral completion," which enhance accuracy, and simultaneous masking and the temporal spread of masking (forward and backward masking), which reduce accuracy. In addition, there is strong evidence that accuracy is enhanced by context, which is present in sentences but lacking in monosyllabic word lists.

DISCUSSION

As discussed in the introduction, the interference produced by spectral alternation appears to be similar to that produced by dichotic alternation (Cherry, 1953). Both paradigms produce a subjective splitting of the speech signal into separate streams of interrupted speech, segregated either by spectral range or by ear of presentation, and the two forms of alternation interfere with intelligibility at similar rates of switching. However, presenting adjacent fragments of the speech signal to different spectral channels rather than to opposite ears of the listener appears to produce greater interference. The adverse effect of spectral alternation was demonstrated in the present study using a conventional method of intelligibility testing. In contrast, dichotic alternation has little effect upon objective intelligibility when listeners are not required to perform the difficult task of shadowing continuous speech (Hopkinson, 1967; Kreul, 1971; Speaks & Trooien, 1974). In conditions in the present study that involved only single bands of filtered speech interrupted by silence, repetition accuracy was considerably lower than that previously observed with broadband sentences (Powers & Wilcox, 1977) and monosyllables (Miller & Licklider, 1950) interrupted by silence at similar rates. However, the greater interference produced by spectral alternation than by dichotic alternation seems to reflect more than a loss of information due to filtering. Preliminary experiments in this laboratory using dichotically alternated sentences and monosyllables subjected continuously to either of the filtering conditions employed in the present study provided little if any evidence of an objective loss in repetition accuracy.

As mentioned earlier, the recovery of intelligibility that resulted from the addition of complementary noise bands to spectrally alternating speech bands can be considered as the net effect of opposing processes of restoration and masking produced by the noise. Due to the spread of masking across different spectral regions, each noise band would be expected to produce some simultaneous masking of the spectrally complementary speech with which it was mixed, but due to the spectral asymmetry of masking (its greater upward spread), interference with intelligibility would more likely be produced by the addition of low-pass noise to high-pass speech. This effect was observed in conditions in the present study that presented only one speech band. Furthermore, due to the temporal spread of masking to the preceding signal (backward masking) and to the following signal (forward masking), bursts of noise occurring in either frequency region would also be expected to produce some interference with temporally adjacent fragments of speech that had the same spectral limits. No loss of intelligibility attributable to the temporal spread of masking was observed in conditions in this study that included only an interpolated noise band. However, when simultaneous low-pass noise was mixed with fragments of high-pass filtered sentences, the further addition of interpolated high-pass noise reduced,

rather than enhanced, intelligibility. It appears that the simultaneous masking produced by the low-pass noise in that condition resulted in a loss of linguistic context sufficient to inhibit phonemic restoration, thereby revealing the otherwise hidden ability of the high-pass noise to function as a forward and backward masker of the high-pass speech fragments.

The temporal spread of masking observed with single speech bands was negated and, under some conditions, substantially reversed when both speech bands were present. As discussed earlier, the net increase in intelligibility produced by adding complementary noise to the alternating speech bands was presumably due in part to phonemic restorations occurring within spectrally limited regions of the speech signal. However, the enhancement of intelligibility was considerably greater than had previously been observed with similar stimuli subjected to periodic interruption occurring simultaneously in all spectral regions. Powers and Wilcox (1977), who used broadband CID sentences interrupted by broadband noise at various rates and relative noise levels, observed a maximum noise-induced increase in repetition accuracy of about 20%. Similarly, Bashford and Warren (1979), who presented listeners with a 1500-Hz band-pass version of the same recording of CID sentences used in the present study, observed a maximum increase of about 17% in repetition accuracy when the filtered sentences were interrupted by spectrally matched noise rather than silence. Much greater improvement was observed in the present study when the same stimuli were subjected to spectral alternation: The addition of high-pass noise alone produced an increase of 29%, and the addition of both noise bands produced an increase of 39%. Furthermore, the addition of complementary noise to spectrally alternating monosyllables also improved intelligibility. In previous studies, interpolated noise was found to produce only masking of monosyllables subjected either to periodic interruption (Dirks & Bower, 1970; Miller & Licklider, 1950) or to dichotic alternation (Hopkinson, 1967; Kreul, 1971).

The greater noise-induced recovery of intelligibility observed under conditions of spectral alternation does not appear to be due to a simple summation of restoration effects occurring within each frequency region. The increase in intelligibility produced by the addition of low-pass noise to the gaps in the interrupted low-pass speech presented alone did not reach significance (see Conditions 5 and 6 in Table 1), nor was there evidence of restoration for isolated monosyllables produced by either noise band (see Conditions 5 and 6 and Conditions 7 and 8). As discussed below, the greater effectiveness of noise in the present study can be attributed to a previously unreported type of intelligibility enhancement operating under conditions of spectral alternation, as shown in Figure 1B.

In previous studies dealing with the effects of interpolated noise upon intelligibility, interruptions removed all spectral components of the speech signal simultaneously, so that the only context directing restoration was provided by speech fragments temporally adjacent to the interrupt-

ing noise. Under the present conditions of spectral alternation, the interrupting noise was mixed with concurrent and temporally appropriate verbal context from a different frequency region, so that a simultaneous spectral restoration or completion could also provide a basis for reconstructive synthesis of missing speech components, which could then enhance the phonemic restorations within each frequency band of speech. The most striking example of this effect was observed with sentences. The enhancement of sentence intelligibility produced by the addition of high-pass noise to gaps in high-pass speech presented alone (Condition 7 vs. Condition 8) was only about 12%, but the increase in intelligibility was about 29% when the high-pass noise bursts were mixed with fragments of low-pass speech (Condition 1 vs. Condition 3). Thus, it appears that replacing portions of speech with spectrally matched noise can permit two forms of perceptual restoration: (1) temporal completion (phonemic restoration) based upon prior and subsequent context within the same frequency region, and (2) spectral completion based upon concurrent context in a different frequency region. This added effect of spectral completion may also have been responsible for occasional reports by the listeners in pilot studies that the addition of complementary noise to the alternating speech bands produced an apparent fusion of the two streams of intermittent speech into a continuous broadband message.

Although the finding of spectral completion of speech appears to be a novel one, somewhat similar effects have recently been reported to occur with tonal stimuli. Plomp and Larkin (cited in Plomp, 1981) found that the continuous removal of several adjacent harmonics from a complex tone had little or no effect upon the timbre of the tone so long as the missing harmonics were replaced by noise of sufficient amplitude to have masked them had they actually been present, and so long as listeners were biased through prior listening to the tone with all harmonics present. Furthermore, Houtgast (1976) found that a "low pitch" similar to that normally evoked by a complex of harmonically related sinusoids (the so-called missing fundamental of a harmonic series) could be produced by a single tone (i.e., a sinusoid of x Hz was found to produce a "subharmonic" pitch of x Hz divided by an integer) when the sinusoid was presented slightly above threshold in broadband noise, and when listeners were primed through prior listening to an appropriate harmonic complex of sinusoids. The perceptual effects of noise in these nonverbal studies indicate that a form of spectral completion can restore spectral components of steady-state complex tones obliterated by noise. The results of the present study indicate that linguistically appropriate spectral completion can also occur with dynamic speech stimuli partially masked by spectrally limited noise.

REFERENCES

- AMERICAN NATIONAL STANDARDS INSTITUTE. (1960). *USA Standard: Method for measurement of monosyllabic word intelligibility* (Reprint S3.2-1960 [R1971]). New York: Author.
- BASHFORD, J. A., JR., & WARREN, R. M. (1979). Perceptual synthesis of deleted phonemes. In J. J. Wolf & D. H. Klatt (Eds.), *Speech communication papers* (pp. 423-426). New York: Acoustical Society of America.
- BASHFORD, J. A., JR., & WARREN, R. M. (1987). Multiple phonemic restorations follow the rules for auditory induction. *Perception & Psychophysics*, **42**, 114-12.
- BREGMAN, A. S., & CAMPBELL, J. (1971). Primary auditory stream segregation and perception of order in rapid sequences of tones. *Journal of Experimental Psychology*, **89**, 244-249.
- BROADBENT, D. E. (1958). *Perception and communication*. London: Pergamon.
- BUKOFZER, M. (1947). *Music in the Baroque era*. New York: W. W. Norton.
- CHERRY, C. E. (1953). Some experiments on the recognition of speech, with one and with two ears. *Journal of the Acoustical Society of America*, **25**, 975-979.
- CHERRY, C. E., & TAYLOR, W. K. (1954). Some further experiments upon the recognition of speech, with one and with two ears. *Journal of the Acoustical Society of America*, **26**, 554-559.
- CHERRY, C. E., & WILEY, R. (1967). Speech communications in very noisy environments. *Nature*, **214**, 1164.
- DAVIS, H., & SILVERMAN, S. R. (1970). *Hearing and deafness*. New York: Holt, Rinehart & Winston.
- DEUTSCH, D. (1972). Octave generalization and tune recognition. *Perception & Psychophysics*, **11**, 411-412.
- DIRKS, D. D., & BOWER, D. (1970). Effect of forward and backward masking on speech intelligibility. *Journal of the Acoustical Society of America*, **47**, 1003-1007.
- DOWLING, W. J. (1973). The perception of interleaved melodies. *Cognitive Psychology*, **5**, 322-337.
- DOWLING, W. J., & HOLLOWAY, A. W. (1977). The perception of melodies distorted by splitting into several octaves: Effects of increasing proximity and melodic contour. *Perception & Psychophysics*, **21**, 60-64.
- HOLLOWAY, C. M. (1970). Passing the strongly voiced components of noisy speech. *Nature*, **226**, 178-179.
- HOPKINSON, N. T. (1967). Combined effects of interruption and interaural alternation on speech intelligibility. *Language & Speech*, **10**, 234-243.
- HOUTGAST, T. (1972). Psychophysical evidence for lateral inhibition in hearing. *Journal of the Acoustical Society of America*, **51**, 1885-1894.
- HOUTGAST, T. (1976). Subharmonic pitches of a pure tone at low S/N ratio. *Journal of the Acoustical Society of America*, **60**, 405-409.
- HUGGINS, A. W. F. (1964). Distortion of the temporal pattern of speech: Interruption and alternation. *Journal of the Acoustical Society of America*, **36**, 1055-1064.
- HUGGINS, A. W. F. (1967). Distortion of the temporal pattern of speech by syllable-tied alternation. *Language & Speech*, **10**, 133-140.
- KREUL, E. J. (1971). Speech intelligibility for interaural alternated speech with and without intervening noise for words and nonsense. *Language & Speech*, **14**, 99-107.
- LAYTON, B. (1975). Differential effects of two nonspeech sounds on phonemic restoration. *Bulletin of the Psychonomic Society*, **6**, 487-490.
- MILLER, G. A., & HEISE, G. A. (1950). The trill threshold. *Journal of the Acoustical Society of America*, **22**, 637-638.
- MILLER, G. A., & LICKLIDER, J. C. R. (1950). The intelligibility of interrupted speech. *Journal of the Acoustical Society of America*, **22**, 167-173.
- OBUSEK, C., & WARREN, R. M. (1973). Relation of the verbal transformation and the phonemic restoration effects. *Cognitive Psychology*, **5**, 97-107.
- ORTMANN, O. (1926). On the melodic relativity of tones. *Psychological Monographs*, **35** (Whole No. 162).
- PISTON, W. (1947). *Counterpoint*. New York: W. W. Norton.
- PLOMP, R. (1981). Perception of sound signals at low signal-to-noise ratios. In D. J. Getty & J. H. Howard, Jr. (Eds.), *Auditory and visual pattern recognition* (pp. 27-35). Hillsdale, NJ: Erlbaum.
- POWERS, G. L., & WILCOX, J. C. (1977). Intelligibility of temporally interrupted speech with and without intervening noise. *Journal of the Acoustical Society of America*, **61**, 195-199.

- SAMUEL, A. G. (1981a). Phonemic restoration: Insights from a new methodology. *Journal of Experimental Psychology: General*, **110**, 474-494.
- SAMUEL, A. G. (1981b). The role of bottom-up confirmation in the phonemic restoration illusion. *Journal of Experimental Psychology: Human Perception & Performance*, **7**, 1124-1131.
- SAMUEL, A. G. (1987). Lexical uniqueness effects on phonemic restoration. *Journal of Memory & Language*, **26**, 36-56.
- SAMUEL, A. G., & RESSLER, W. H. (1986). Attention within auditory word perception: Insights from the phonemic restoration illusion. *Journal of Experimental Psychology: Human Perception & Performance*, **12**, 70-79.
- SASAKI, T. (1980). Sound restoration and temporal localization of noise in speech and music sounds. *Tohoku Psychologica Folia*, **39**, 79-88.
- SCHUBERT, E. D., & PARKER, C. D. (1955). Addition to Cherry's finding on switching speech between the two ears. *Journal of the Acoustical Society of America*, **27**, 792-794.
- SILVERMAN, S. R., & HIRSH, I. J. (1955). Problems related to the use of speech in clinical audiometry. *Annals of Otology, Rhinology, & Laryngology*, **64**, 1234-1245.
- SPEAKS, C., & TROOIJEN, T. T. (1974). Interaural alternation and speech intelligibility. *Journal of the Acoustical Society of America*, **56**, 640-644.
- VAN NOORDEN, L. P. A. S. (1975). *Temporal coherence in the perception of tone sequences*. Doctoral dissertation, Eindhoven University of Technology, The Netherlands.
- VERSCHUURE, J. (1978). *Auditory excitation patterns*. Doctoral dissertation, Erasmus University, Rotterdam. Delft: W. D. Meinema.
- VERSCHUURE, J., & BROCAAR, M. P. (1983). Intelligibility of interrupted meaningful and nonsense speech with and without intervening noise. *Perception & Psychophysics*, **33**, 232-240.
- WARREN, R. M. (1970). Perceptual restoration of missing speech sounds. *Science*, **167**, 392-393.
- WARREN, R. M. (1984). Perceptual restoration of obliterated sounds. *Psychological Bulletin*, **96**, 371-383.
- WARREN, R. M., & OBUSEK, C. J. (1971). Speech perception and phonemic restorations. *Perception & Psychophysics*, **9**, 358-362.
- WARREN, R. M., OBUSEK, C. J., & ACKROFF, J. M. (1972). Auditory induction: Perceptual synthesis of absent sounds. *Science*, **176**, 1149-1151.
- WARREN, R. M., & SHERMAN, G. L. (1974). Phonemic restorations based on subsequent context. *Perception & Psychophysics*, **16**, 150-156.
- WINGFIELD, A., & WHEALE, J. L. (1975). Word rate and intelligibility of alternated speech. *Perception & Psychophysics*, **18**, 317-320.

NOTES

1. It was found in preliminary experiments that spectral alternation produced greater interference with intelligibility when the filtering cutoffs were adjusted to produce minimal spectral overlap between the two speech bands. When a spectral overlap between the two bands was present, a continuous narrowband version of the message was also heard corresponding to the frequency overlap. In order to prevent this interesting but unwanted effect from occurring in the present experiment, a spectral gap was introduced, with the filtering cutoffs chosen to produce a 40-dB attenuation of each speech band at their point of spectral overlap.

2. Filtering cutoffs were chosen on the basis of pilot work indicating that both the 1100-Hz low-pass and the 1700-Hz high-pass condition permitted nearly perfect repetition accuracy for CID sentences presented without interruption. Unfortunately, with the less redundant PB words, and under conditions of interruption, the high-pass band proved to be of somewhat higher intelligibility.

3. Briefly, Newman-Keuls analysis of the sentence data indicated that Stimulus Conditions 5, 6, 7, 9, 10, and 11 produced equivalent repetition accuracy, as did Conditions 5 and 12 and Conditions 1 and 2. All other contrasts were significant at the .05 level or better. Analysis of PB word data indicated that repetition accuracy was equivalent in Conditions 5, 6, 9, 10, 11, and 12. Performance was also equivalent between Conditions 8 and 7, 8 and 11, 8 and 12, 1 and 2, and 3 and 4. The remaining contrasts were significant at the .05 level or better.

(Manuscript received November 17, 1986;
revision accepted for publication April 25, 1987.)